Dynamic index finger gesture video dataset for mobile interaction

C. Arslan, J. Martinet, I. M. Bilasco
University of Lille, CRIStAL, CNRS, Lille, France
cagan.arslan@univ-lille.fr, jean.martinet@univ-lille.fr, marius.bilasco@univ-lille.fr

Université de Lille

1. Introduction

Video analysis is an important aspect of multimedia data description. A basic task in video analysis is the extraction of optical flow, that help understanding individual region motion in the video stream. We address a specific scenario of one-finger gesture in the context of mobile interaction. In this scenario, a user interacts with a mobile phone using the index finger on phone's back camera equipped with a wide angle lens. The ability to precisely capture index finger gestures opens interaction opportunities for daily tasks in mobile contexts, such as controlling widgets, sending messages, playing video games and so on. We present a new index finger dataset including 6 gestures together with a baseline classification algorithm dedicated to this dataset. The 6 dynamic gesture classes are swipe left, swipe right, swipe up, swipe down, tap on the lens and a counter-clockwise circle (Fig. 2).

2. Experimental Setup

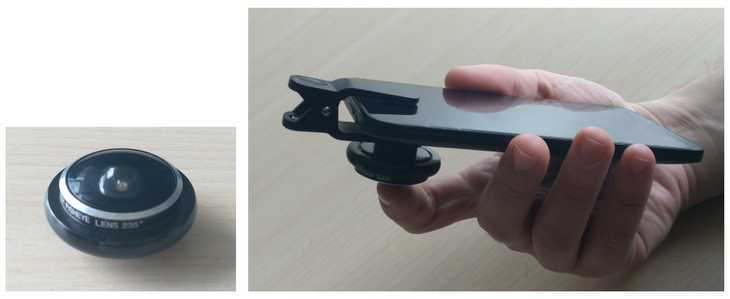

Figure 1: Left: Wide-angle lens.
Right: Interacting with the lens.

The gestures are performed 9 times by 14 users (2 left handed), and captured with a low-cost wide angle (235°) lens. As a result, 756 video sequences with a resolution of 320x240 pixels are obtained. Sequences were taken over different backgrounds and in uncontrolled lighting conditions.

3. Dataset

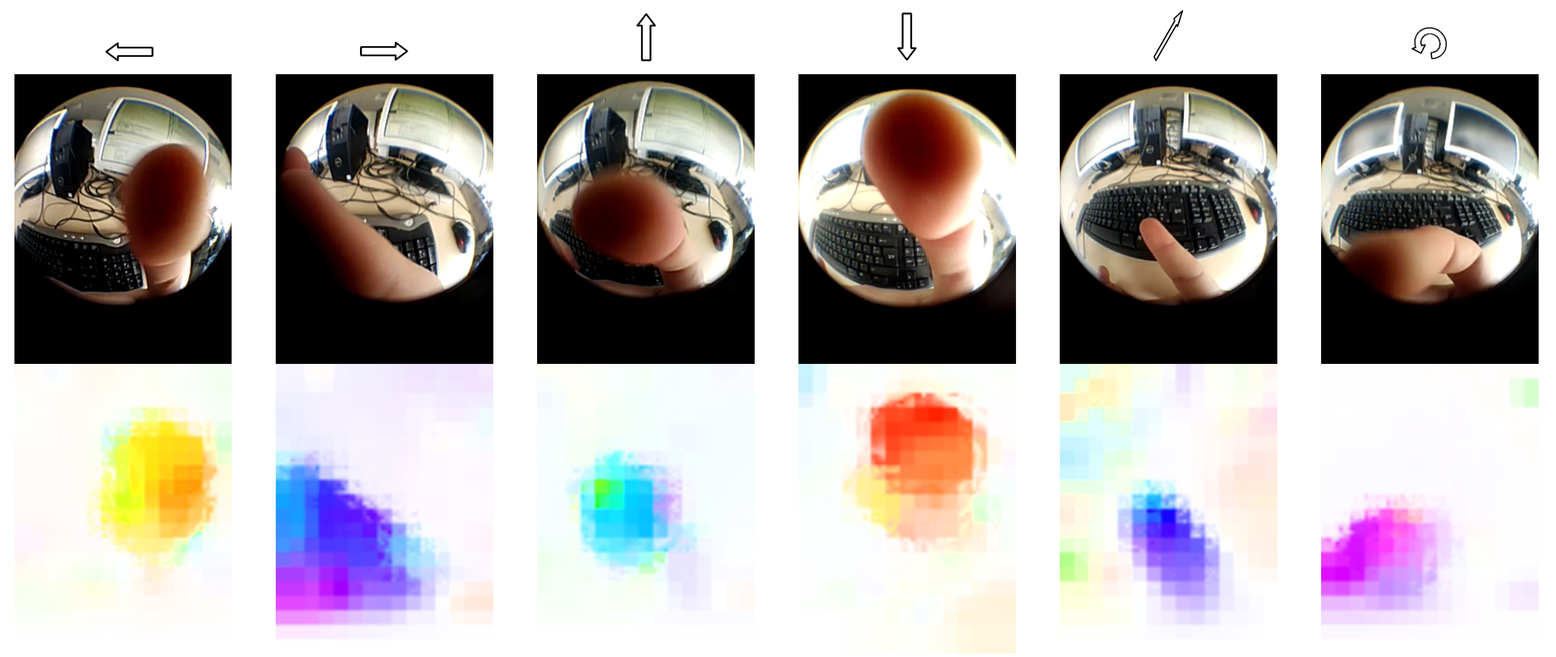

Figure 2: 6 gestures in the dataset. (Top row) Example frames from the start of each gesture. (Bottom row) Color representation of the optical flow.

4. Challenges

- The wide angle lens introduces shape deformation and occasional lens glare.
- Moving the finger causes small changes in the orientation of the device (hand tremor). This adds a global motion to the image sequence at the time of the gesture.
- Left, right, up and down, the gestures are almost diagonal because of the orientation of the index finger behind the device.

5. Baseline Method

In order to match the computationally constrained scenario of using a smartphone we keep our pipeline simplistic. For a given gesture video sequence, a dense optical flow is calculated in every frame. In order to filter out the hand tremor, we discard the flow vectors with a magnitude less than a threshold ($t_{magnitude}$). If the ratio of remaining vectors to image size is smaller than a threshold $t_{vectors}$, the frame is discarded as it is not sufficiently descriptive. An 8-bin histogram of directions is constructed to represent distribution of the optical flow in each frame. if a direction in the histogram is represented by less than a percentage ($t_{directions}$) of the non-zero pixels, that bin is cleared. Hence, at each time instant, we have a filtered histogram of the optical flow. To represent the dynamic gesture, we construct a feature vector that is the average histogram corresponding to the image sequence.

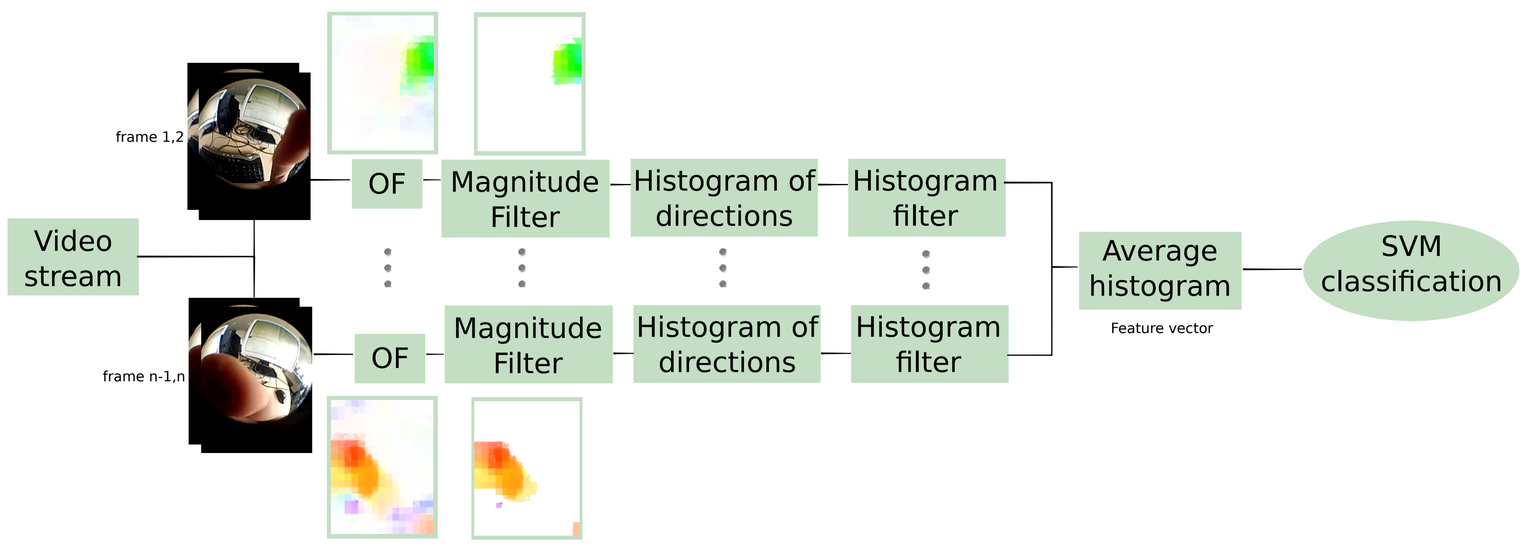

Figure 3: Baseline method

6. Results

For the feature extraction parameters we obtained the best results by choosing $t_{magnitude} = 5$, $t_{vectors} = 0.01$ and $t_{directions} = 0.05$. The average 8-bin histogram vectors is then fed to an SVM with RBF kernel. The average accuracy was obtained by using leave-one-user-out approach. A grid search was used to find the optimal RBF kernel parameters C and γ that give the best average accuracy. In order to demonstrate the need for magnitude filtering, we compared the performance of our proposed method with unfiltered averaged descriptors of histogram of optical flow and of histogram of oriented optical flow.

	Accuracy
HOOF	%63.8889
HOF	%61
Baseline Method	%73.1

Table 1: Accuracies on the dataset

	left	right	up	down	tap	circular
left	0.75	0.01	0	0.02	0.05	0.17
right	0.01	0.83	0	0	0.16	0.01
up	0	0.03	0.80	0.02	0.09	0.06
down	0	0	0.12	0.73	0.14	0.11
tap	0.03	0.19	0.12	0.06	0.56	0.05
circular	0.07	0	0.10	0.04	0.12	0.67

Table 2: Confusion matrix

6. Conclusions

- The proposed dataset brings new interactions opportunities as well as new challenges to deal with such as the ability to deal with index finger gesture in mobile context using back camera, the variations in terms of index finger dynamic, unconstrained usage of the mobile dynamic and challenging backgrounds.
- Captured with a smartphone camera equipped with a wide-angle lens carries distinctive properties such as a non linear field of view and hand tremor that are ready to be exploited.
- As feature work, the dataset can be used for shape based methods such as tracking fingertip position for gesture recognition.